Title:	The POLar SysTem Analysis Package Version 1 Specifications
Author(s):	Roberts, Andrew Frank
Intended for:	Report
Issued:	2021-01-20 (rev.1)

The POLar SysTem Analysis Package

Version 1 Specifications

Andrew Roberts
T-3, Los Alamos National Laboratory

January 20, 2021

Contents

1 Scope

The Polar System Analysis Package (POLeSTAr) is a modern data fusion tool being designed to benchmark, evaluate and compare the polar physics and biogeochemistry of Earth System Models (ESMs). It is being written in Python to make use of libraries including Xarray, Matplotlib, SciPy, NumPy and PyTorch within algorithms developed specifically for polar system modeling. Initial POLeSTAr algorithms will be adapted from the MATLAB Ridgepack package that was designed to analyze the Model for Prediction Across Scales (MPAS) sea ice component of the Energy Exascale Earth System Model (E3SM) and sea ice model CICE within the Community Earth System Model (CESM). POLeSTAr Version 1 will adopt much of the data fusion libraries developed in Ridgepack and convert them to Python, as well as offering new functionality. POLeSTAr will be applicable across the range of component models within E3SM and CESM listed in Table 1. The primary focus of POLeSTAr V1 is on analysis and benchmarking of coupled interactions of CICE, MPAS-SeaIce, and the Discrete Element Model of Sea Ice (DEMSI) codes summarized in Table 2 with other component models in E3SM or CESM. The physical and biogeochemical components of these sea ice models are summarized in Figures 1 and 2, each using different dynamical representations to simulate frozen ocean. POLeSTAr V1 will be interoperable across both Lagrangian paths and Eulerian structured and unstructured meshes (Figure 3). Later POLeSTAr versions are anticipated to guarantee functionality across a broad set of Coupled Model Inter-comparison Project (CMIP) models. However, that requires entrainment of the international earth system modeling community for which POLeSTAr V1 is intended as a proof-of-concept. POLeSTAr will be published as a Conda package and will reside in a GitHub repository. An object-oriented code design is intended to make the package easily extensible.

| Coupled Model	Sea Ice	Ocean	Atmosphere	Land	Runoff
E3SM V3	MPAS-SeaIce (DEMSI)	MPAS-Ocean	EAM	ELM	MOSART
CESM V3	CICE	MOM6	CAM	CLM	MOSART

Table 1: Coupled models and their components for which POLeSTAr V1 is expected to function. Components of the Energy Exascale Earth System Model (E3SM) include MPAS-SeaIce and MPAS-Ocean, as well as the E3SM atmosphere (EAM) and land (ELM) models. The Community Earth System Model (CESM) is being readied for production with CICE coupled to the coupled Modular Ocean Model (MOM6) using the Community Atmosphere (CAM) and Land (CLM) models. Both E3SM and CESM use the Model for Scale Adaptive River Transport (MOSART). Infrastructure for MPAS-SeaIce and MPAS-Ocean will readily adapt to MPAS-Atmosphere should CESM include that as an option in the coming years. The Discrete Element Model for Sea Ice (DEMSI) is expected to become available for coupled simulations in E3SM after POLeSTAr V1 is released.

| Sea Ice Model[†]	Version	Domain	Sea Ice DyCore[§]	
DEMSI	V1	Arctic	LAMMPS	LD
MPAS-SeaIce	V2	global	MPAS	EU
CICE	V6	global	native	ES

Table 2: Sea ice models for which POLeSTAr V1 is designed to work and for which their coupled interactions are the primary focus. These component models are used with or are planned to work with one of E3SM or CESM (see Table 1). † Codes - CICE Consortium sea ice model; MPAS-SeaIce - Sea ice component of the Model for Prediction Across Scales; DEMSI: Discrete Element Model of Sea Ice; § Dynamical Cores - MPAS: Model for the Prediction Across Scales; LAMMPS: Large-scale Atomic/Molecular Massively Parallel Simulator; *native* refers to the CICE dynamical core that does not access a broader dynamics framework. Key: L - Lagrangian DyCore using the discrete element method (D); E - Eulerian DyCore on either a structured (S) or unstructured (U) mesh (see Figure 3); Each model uses or is being configured to use CICE Consortium sub-grid scale Icepack saline ice physics and biogeochemistry.

Figure 1: Components of a sea ice dynamical core (DyCore) for Lagrangian (*L*: DEMSI) and Eulerian (*E*: MPAS-SeaIce and CICE) classes of model summarized in Table 2 for which POLeSTAr will accommodate suitable data and coordinate structures for model inter-comparison within the data fusion core detailed in Figure 3. Icepack represents sub-grid scale physics and biogeochemistry as a submodule, expanded in Figure 2. The dynamical cores also include infrastructure for configuring and running the model and providing output (not shown), such as configuring multiple model options including three methods for simulating internal ice stress within CICE (VP, EVP and EAP).
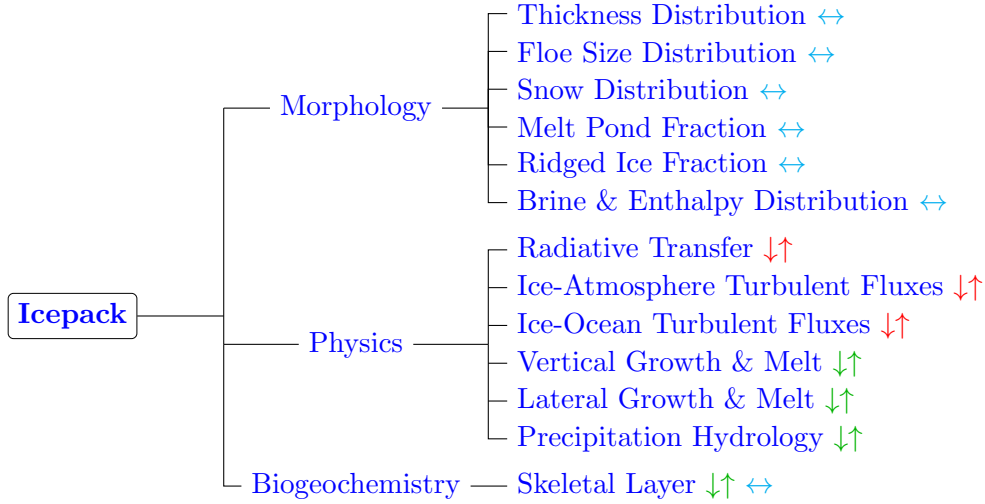

Figure 2: Icepack is the sea ice physics and biogeochemistry column package for which POLeSTAr V1 will primarily focus capabilities on understanding direct and indirect coupling of these components with ocean, atmosphere and land models in E3SM and CESM. Arrows indicate energy (↓↑) and mass (↓↑) flux exchange with the ocean and atmosphere, as well as horizontal advection (↔) using a dynamical core with Icepack, such as CICE.
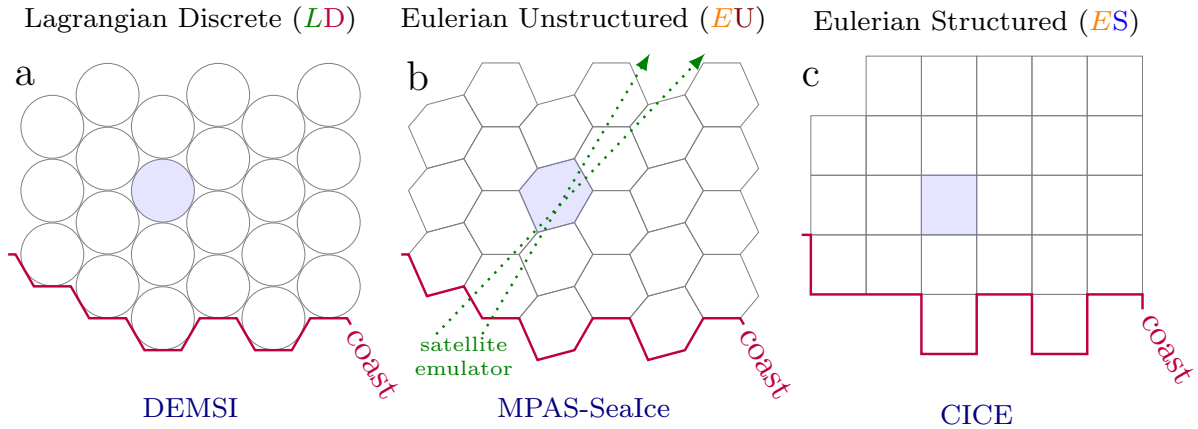
Figure 3: Schematic of mesh types corresponding to sea ice dynamical cores in Table 2 for *L*D, *E*U, and *E*S. Sub-grid scale sea ice physics and biogeochemistry in Figure 2 are represented statistically within elements or cells (blue shading) for (a) DEMSI, (b) MPAS-SeaIce, and (c) CICE. Circular discrete elements in (a) move with the pack, whereas the mesh is fixed in space for (b) and (c). Green tracks in (b) represent satellite ground passes as one example of the emulator capability of the data fusion core (DFCore). The mesh in (c) is both structured and quadrilateral. POLeSTAr's data structures and coordinate objects ease interoperability between all three mesh types, and are also applicable to other components of earth system models.

2 Package Structure

POLeSTAr is to be organized into three main components. Most important of these is the software library at the heart of POLeSTAr's data fusion capabilities, named the Data Fusion Core (DFCore). The DFCore is designed to be extensible, and in Version 1 will have six main components focused on ingesting, interlacing, and writing data using data structures, emulating satellite and in-situ observations in models, publication quality graphics and animation, as well as signal processing applicable across a broad range of physics applications. Figure 4 summarizes these components, and includes a machine learning element expected to be the focus of future versions of the code. The second core POLeSTAr element is the data area, for which we plan to allow model output and observations to be downloaded on-call during code execution. Data will be categorized into Eulerian, Lagrangian, and emulation divisions for models and measurements alike. POLeSTAr will initially prioritize emulation and inter-comparison between models and observational datasets listed in Table 3. In addition, pre-calculated metrics from CESM and E3SM CMIP6 submissions will be included as standard benchmarking data. The third fundamental POLeSTAr component of POLeSTAr is the operating area for users, for which *cases* may be created akin to simulation cases within the Common Infrastructure for Modeling the Earth (CIME). The case area is designed to allow models and their data to be flexibly analyzed in isolation or in concert with the use of advanced metrics to indicate forward model bias and skill. Further to these three core elements, the remaining directories illustrated in Figure 4 will house documentation, Conda package information, and a contributor list, as well as testing scripts.
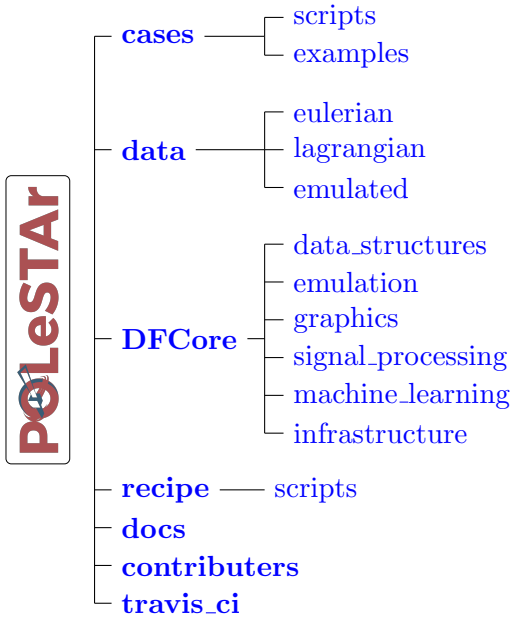
Figure 4: POLeSTAr directory structure down to the third level, indicating composition of the package. There is a **cases** area housing high-level single- and multi-model processing, benchmarking, and analysis scripts that utilize structured and unstructured model and observational **data** (Tables 1, 2, and 3), and utilize core data fusion libraries in (**DFCore**). Observational and model data is divided into E - **eulerian**, L - **lagrangian**, and O - **emulated** components corresponding to Table 3 (measurements) and Figure 3 (model). The data fusion core has a **data_structures** library for placing model and observational data into four main data structures and assigning any of three coordinate objects described in Section 3 to the data. Pre- and post-processed satellite **emulation** for ICESat and ICESat-2 will be included in POLeSTAr V1 (refer to Figure 3 satellite tracks). The package will include publication quality **graphics** and sophisticated **signal_processing** adopted from Ridgepack (see section 1). The **machine_learning** area of POLeSTAr is under consideration, but will not be populated in POLeSTAr V1. The **DFCore** will include debugging tools located in the **infrastructure** area. In addition, there is a Conda **recipe** area, **doc**umentation and **contributors** list, as well testing with **travis_ci**.

3 Data Fusion Core

Data fusion between forward-integrated earth system models and observations requires carefully designed data structures. POLeSTAr V1 will organize internal information around four types of extensible structures within DFCore: a climate and forecast (CF) structure compatible with input and output of the NetCDF CF convention, Keyhole Markup and GeoJSON structures compatible with geographical information systems and Google Earth, and finally a complex data structure for ease in multi-dimensional time series statistics, filtering, and machine learning. Within these structures, three supported coordinate objects will provide maximum flexibility for the purpose: 1) rotated cartesian and spherical coordinates, compatible with structured and unstructured Eulerian meshes, Lagrangian traces, as well as point, vector

Diagnostic	Observations	Duration
Sea Ice State		
Concentration	*E* NOAA Climate Data Record (*Meier et al.*, 2014)	1979-
Landfast Extent	*E* Bureau of Ocean Energy Management Climatology[§]	1996-
Drift	*E* Polar Pathfinder Drift (*Tschudi et al.*, 2016)	1978-
	L International Arctic Buoy Program[†]	1980-
Freeboard	*O* ICESat (*Yi and Zwally*, 2010)	2003-2008
	O IceBridge Freeboard & Snow depth (*Kurtz et al.*, 2015)	2009-2019
	O CryoSat-2 (*Tilling et al.*, 2015)	2010-
	O ICESat-2 (*Markus et al.*, 2017)	2018-
Draft	*O* U.S. Navy and Royal Navy (*NSIDC*, 2006)	1960-2005
Polar Ocean		
Mixing	*L* Ice Tethered Profilers (Timmermans[§])	2004-
Ice-Ocean Flux	*L* Ocean Flux Buoys (*Shaw et al.*, 2008)	2002-2017
	L MOSAiC Ice-Ocean Flux Buoys (Stanton[§])	2019-2020
Atmospheric Flux	*L* SHEBA Flux Tower Data (*Persson*, 2002)	1997-1998
	L MOSAiC Atmospheric Flux (when available)	2019-2020
Tides	*E* TPXO Arctic and Antarctic Atlas[‡]	

Table 3: Baseline observational datasets to be given scripting priority for model evaluation in the first version of POLeSTAr. Key: *E* - Eulerian mapping; *L* - Lagrangian tracking; and *O* - Observational emulation. References: NOAA - National Oceanic and Atmospheric Administration; SHEBA - Surface Heat Balance of the Arctic *in situ* data; MOSAiC - Multidisciplinary drifting Observatory for the Study of Arctic Climate *in situ* data; † - http://iabp.apl.washington.edu; § - Los Alamos National Laboratory collaborators; ‡ - https://www.tpxo.net

and polygonal geographical data; 2) a rotated stereographic projection with Cartesian coordinates that can be made azimuthal to any point on Earth following *Snyder* (1983), and; 3) time objects compatible with CF-convention and most CMIP-model calendars using one of either Gregorian, proleptic Gregorian, or perpetual no leap (365-day) serial time referenced to a past date. This approach to organizing information accommodates *cmorized* versions of coupled model output, raw model and observational series on native meshes or Lagrangian paths, as well as model-observation fusion, including for ICESat and ICESat-2 satellite emulators as illustrated in Figure 3b. Conversion routines for ingesting data to these formats will follow an easily-duplicated format and flag possible discrepancies with base datasets.

The main point of careful data organization is to facilitate inference and discovery across large model and observational datasets. The planned data structures and objects will easily facilitate high-level signal processing and statistical inference in concert with observational emulation in DFCore, including: 1) rotary wavelet and spectral analysis and filtering for tidal and high-frequency (inertial resolving) ice, ocean and atmospheric coupled model output and coherence testing against in-situ observations; 2) traditional skill and bias analysis as summarized with Taylor diagrams and bias tables; 3) model-observation compatibility analysis, including comparing highly autocorrelated datasets in line with current practice (e.g. *Amrhein et al.*, 2019; *Bishop*, 2019); 4) Machine learning to categorize model and observational data, assess their differences, and to help generate optimal parameter spaces for models, and; 5) crisp, clear and easily created informational graphics to visualize and cross-reference data on native meshes and Lagrangian paths. For this, the POLeSTAr prototype will adapt low-level methods from Ridgepack with data-specific color-density (e.g. *Zeller and Rogers*, 2020) to visualize scalar and vector fields and their interactions using images and animations appropriate for publication and near-real time post-simulation analysis.

4   Community Engagement

POLeSTAR V1 is intended to address limitations of existing community tools for benchmarking the polar realms of coupled earth system models applicable to existing and emerging ocean, sea ice, and atmospheric dynamical cores, physics, and biogeochemistry. The CICE Consortium will be used as the collaboration forum for POLeSTAr development, extending awareness of the package beyond personnel from Los Alamos National Laboratory (LANL) and the National Center for Atmospheric Research (NCAR) who are the primary developers. POLeSTAR will be publicly available through GitHub, adhering to Coordinated Model Evaluation Capabilities (CMEC) standards. Documentation is critical for community engagement and is planned to be version-specific using Doxygen. Code will be published with a DOI number, referenced in upcoming publications, including one aimed at community data fusion capabilities as well as a LANL Technical Report on POLeSTAr V1. The creation and capabilities of POLeSTAr will be announced at workshops and conferences to help entrain effort for future development of this tool.

5   Acknowledgements

6   References

Amrhein, V., S. Greenland, and B. Mcshane (2019), Retire statistical significance, *Nature*, *567*, 305–306, doi:10.1038/d41586-019-00857-9.

Bishop, D. (2019), Rein in the four horsemen of irreproducibility, *Nature*, *568*(7753), 435–435, doi:10.1038/d41586-019-01307-2.

Kurtz, N., M. S. Studinger, J. Harbeck, V. Onana, and D. Yi (2015), *IceBridge L4 Sea Ice Freeboard, Snow Depth, and Thickness, Version 1*, Digital Media, NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado USA., doi:10.5067/G519SHCKWQV6.

Markus, T., T. Neumann, A. Martino, W. Abdalati, K. Brunt, B. Csatho, S. Farrell, H. Fricker, A. Gardner, D. Hardin, M. Jasinski, R. Kwok, L. Magruder, D. Lubin, S. Luthcke, J. Morison, R. Nelson, A. Neuenschwander, S. Palm, S. Popescu, C. Shum, B. E. Schutz, B. Smith, Y. Yang, and J. Zwally (2017), The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation, *Remote Sens. Environ.*, *190*, 260–273, doi:10.1016/j.rse.2016.12.029.

Meier, W. N., G. Peng, D. J. Scott, and M. H. Savoie (2014), Verification of a new NOAA/NSIDC passive microwave sea-ice concentration climate record, *Polar Res.*, *33*, 1–22, doi:10.3402/polar.v33.21004.

NSIDC (2006), *Submarine Upward Looking Sonar Ice Draft Profile Data and Statistics, Version 1*, Digital Media, National Snow and Ice Data Center, Boulder, Colorado, USA.

Persson, P. O. G. (2002), Measurements near the Atmospheric Surface Flux Group tower at SHEBA: Near-surface conditions and surface energy budget, *J. Geophys. Res.*, *107*(C10), 1–35, doi:10.1029/2000JC000705.

Shaw, W. J., T. P. Stanton, M. G. McPhee, and T. Kikuchi (2008), Estimates of surface roughness length in heterogeneous under-ice boundary layers, *J. Geophys. Res. Ocean.*, *113*, C08030, doi:10.1029/2007JC004550.

Snyder, J. P. (1983), Map Projections Used by the U.S. Geological Survey, *Tech. rep.*

Tilling, R. L., A. Ridout, A. Shepherd, and D. J. Wingham (2015), Increased Arctic sea ice volume after anomalously low melting in 2013, *Nat. Geosci.*, *8*(8), 643–646, doi:10.1038/ngeo2489.

Tschudi, M., C. Fowler, J. Maslanik, J. S. Stewart, and W. Meier (2016), *Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors, Version 3*, Digital Media, NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado, USA., doi:10.5067/O57VAIT2AYYY.

Yi, D., and J. Zwally (2010), *Arctic Sea Ice Freeboard and Thickness*, Digital Media, National Snow and Ice Data Center, Boulder, Colorado, USA.

Zeller, S., and D. Rogers (2020), Visualizing Science: How Color Determines What We See, *Eos Trans. AGU*, *101*, doi:10.1029/2020EO144330.